Databricks

Exam Questions Databricks-Generative-AI-Engineer-Associate

Databricks Certified Generative AI Engineer Associate

NEW QUESTION 1
A Generative Al Engineer is tasked with improving the RAG quality by addressing its inflammatory outputs.
Which action would be most effective in mitigating the problem of offensive text outputs?

A. Increase the frequency of upstream data updates
B. Inform the user of the expected RAG behavior
C. Restrict access to the data sources to a limited number of users
D. Curate upstream data properly that includes manual review before it is fed into the RAG system

Answer: D

NEW QUESTION 2
A Generative Al Engineer wants their (inetuned LLMs in their prod Databncks workspace available for testing in their dev workspace as well. All of their workspaces are Unity Catalog enabled and they are currently logging their models into the Model Registry in MLflow.
What is the most cost-effective and secure option for the Generative Al Engineer to accomplish their gAi?

A. Use an external model registry which can be accessed from all workspaces
B. Setup a script to export the model from prod and import it to dev.
C. Setup a duplicate training pipeline in dev, so that an identical model is available in dev.
D. Use MLflow to log the model directly into Unity Catalog, and enable READ access in the dev workspace to the model.

Answer: D

NEW QUESTION 3
A Generative Al Engineer at an automotive company would like to build a question- answering chatbot for customers to inquire about their vehicles. They have a database containing various documents of different vehicle makes, their hardware parts, and common maintenance information.
Which of the following components will NOT be useful in building such a chatbot?

A. Response-generating LLM
B. Invite users to submit long, rather than concise, questions
C. Vector database
D. Embedding model

Answer: B

NEW QUESTION 4
A Generative Al Engineer is tasked with developing an application that is based on an open source large language model (LLM). They need a foundation LLM with a large context window.
Which model fits this need?

A. DistilBERT
B. MPT-30B
C. Llama2-70B
D. DBRX

Answer: C

NEW QUESTION 5
A Generative Al Engineer is creating an LLM-based application. The documents for its retriever have been chunked to a maximum of 512 tokens each. The Generative Al Engineer knows that cost and latency are more important than quality for this application. They have several context length levels to choose from.
Which will fulfill their need?

A. context length 514; smallest model is 0.44GB and embedding dimension 768
B. context length 2048: smallest model is 11GB and embedding dimension 2560
C. context length 32768: smallest model is 14GB and embedding dimension 4096
D. context length 512: smallest model is 0.13GB and embedding dimension 384

Answer: D

NEW QUESTION 6
A Generative Al Engineer has created a RAG application which can help employees retrieve answers from an internal knowledge base, such as Confluence pages or Google Drive. The prototype application is now working with some positive feedback from internal company testers. Now the Generative Al Engineer wants to formally evaluate the system??s performance and understand where to focus their efforts to further improve the system.
How should the Generative Al Engineer evaluate the system?

A. Use cosine similarity score to comprehensively evaluate the quality of the final generated answers.
B. Curate a dataset that can test the retrieval and generation components of the system separatel
C. Use MLflow??s built in evaluation metrics to perform the evaluation on the retrieval and generation components.
D. Benchmark multiple LLMs with the same data and pick the best LLM for the job.
E. Use an LLM-as-a-judge to evaluate the quality of the final answers generated.

Answer: B

NEW QUESTION 7
A Generative Al Engineer interfaces with an LLM with prompt/response behavior that has been trained on customer calls inquiring about product availability. The

LLM is designed to output ??In Stock?? if the product is available or only the term ??Out of Stock?? if not.
Which prompt will work to allow the engineer to respond to call classification labels correctly?

A. Respond with ??In Stock?? if the customer asks for a product.
B. You will be given a customer call transcript where the customer asks about product availabilit
C. The outputs are either ??In Stock?? or ??Out of Stock??. Format the output in JSON, for example: {??call_id??: ??123??, ??label??: ??In Stock??}.
D. Respond with ??Out of Stock?? if the customer asks for a product.
E. You will be given a customer call transcript where the customer inquires about product availabilit
F. Respond with ??In Stock?? if the product is available or ??Out of Stock?? if not.

Answer: B

NEW QUESTION 8
A Generative AI Engineer received the following business requirements for an external chatbot.
The chatbot needs to know what types of questions the user asks and routes to appropriate models to answer the questions. For example, the user might ask about upcoming event details. Another user might ask about purchasing tickets for a particular event.
What is an ideal workflow for such a chatbot?

A. The chatbot should only look at previous event information
B. There should be two different chatbots handling different types of user queries.
C. The chatbot should be implemented as a multi-step LLM workflo
D. First, identify the type of question asked, then route the question to the appropriate mode
E. If it??s an upcoming event question, send the query to a text-to-SQL mode
F. If it??s about ticket purchasing, the customer should be redirected to a payment platform.
G. The chatbot should only process payments

Answer: C

NEW QUESTION 9
A Generative AI Engineer is building a Generative AI system that suggests the best matched employee team member to newly scoped projects. The team member is selected from a very large team. The match should be based upon project date availability and how well their employee profile matches the project scope. Both the employee profile and project scope are unstructured text.
How should the Generative AI Engineer architect their system?

A. Create a tool for finding available team members given project date
B. Embed all project scopes into a vector store, perform a retrieval using team member profiles to find the best team member.
C. Create a tool for finding team member availability given project dates, and another tool that uses an LLM to extract keywords from project scope
D. Iterate through available team members?? profiles and perform keyword matching to find the best available team member.
E. Create a tool to find available team members given project date
F. Create a second tool that can calculate a similarity score for a combination of team member profile and the project scop
G. Iterate through the team members and rank by best score to select a team member.
H. Create a tool for finding available team members given project date
I. Embed team profiles into a vector store and use the project scope and filtering to perform retrieval to find the available best matched team members.

Answer: D

NEW QUESTION 10
A Generative AI Engineer has built an LLM-based system that will automatically translate user text between two languages. They now want to benchmark multiple LLM's on this task and pick the best one. They have an evaluation set with known high quality translation examples. They want to evaluate each LLM using the evaluation set with a performant metric.
Which metric should they choose for this evaluation?

A. ROUGE metric
B. BLEU metric
C. NDCG metric
D. RECALL metric

Answer: B

NEW QUESTION 10
A Generative AI Engineer is deciding between using LSH (Locality Sensitive Hashing) and HNSW (Hierarchical Navigable Small World) for indexing their vector database Their top priority is semantic accuracy
Which approach should the Generative AI Engineer use to evaluate these two techniques?

A. Compare the cosine similarities of the embeddings of returned results against those of a representative sample of test inputs
B. Compare the Bilingual Evaluation Understudy (BLEU) scores of returned results for a representative sample of test inputs
C. Compare the Recall-Onented-Understudy for Gistmg Evaluation (ROUGE) scores of returned results for a representative sample of test inputs
D. Compare the Levenshtein distances of returned results against a representative sample of test inputs

Answer: A

NEW QUESTION 14
A Generative AI Engineer just deployed an LLM application at a digital marketing company that assists with answering customer service inquiries.
Which metric should they monitor for their customer service LLM application in production?

A. Number of customer inquiries processed per unit of time
B. Energy usage per query
C. Final perplexity scores for the training of the model

D. HuggingFace Leaderboard values for the base LLM

Answer: A

NEW QUESTION 16
A Generative AI Engineer is testing a simple prompt template in LangChain using the code below, but is getting an error.

```
from langchain.chains import LLMChain
from langchain_community.llms import OpenAI
from langchain_core.prompts import PromptTemplate

prompt_template = "Tell me a {adjective} joke"

prompt = PromptTemplate(
    input_variables=["adjective"],
    template=prompt_template
)

llm = LLMChain(prompt=prompt)
llm.generate([{"adjective": "funny"}])
```

Assuming the API key was properly defined, what change does the Generative AI Engineer need to make to fix their chain?
A)

```
prompt_template = "Tell me a {adjective} joke"

prompt = PromptTemplate(
    input_variables=["adjective"],
    template=prompt_template
)

llm = LLMChain(prompt=prompt)
llm.generate("funny")
```

B)

```
prompt_template = "Tell me a {adjective} joke"

prompt = PromptTemplate(
    input_variables=["adjective"],
    template=prompt_template
)

llm = LLMChain(prompt=prompt.format("funny"))
llm.generate()
```

C)

```
prompt_template = "Tell me a {adjective} joke"

prompt = PromptTemplate(
     input_variables=["adjective"],
     template=prompt_template
     llm=OpenAI()
)

llm = LLMChain(prompt=prompt)
llm.generate([{"adjective": "funny"}])
```

D)

```
prompt_template = "Tell me a {adjective} joke"

prompt = PromptTemplate(
     input_variables=["adjective"],
     template=prompt_template
)

llm = LLMChain(llm=OpenAI(), prompt=prompt)
llm.generate([{"adjective": "funny"}])
```

A. Option A
B. Option B
C. Option C
D. Option D

Answer: C

NEW QUESTION 19
.....

Thank You for Trying Our Product

We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questons and Answers in PDF Format

Databricks-Generative-AI-Engineer-Associate Practice Exam Features:

* Databricks-Generative-AI-Engineer-Associate Questions and Answers Updated Frequently

* Databricks-Generative-AI-Engineer-Associate Practice Questions Verified by Expert Senior Certified Staff

* Databricks-Generative-AI-Engineer-Associate Most Realistic Questions that Guarantee you a Pass on Your FirstTry

* Databricks-Generative-AI-Engineer-Associate Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year